

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-114697

(43)Date of publication of application : 18.04.2003

(51)Int.Cl. G10L 15/20  
B60R 11/02  
G01C 21/00  
G10L 15/00  
G10L 15/28  
G10L 21/02  
// G06F 3/16

(21)Application number : 2001-307200

(71)Applicant : ALPINE ELECTRONICS INC

(22)Date of filing : 03.10.2001

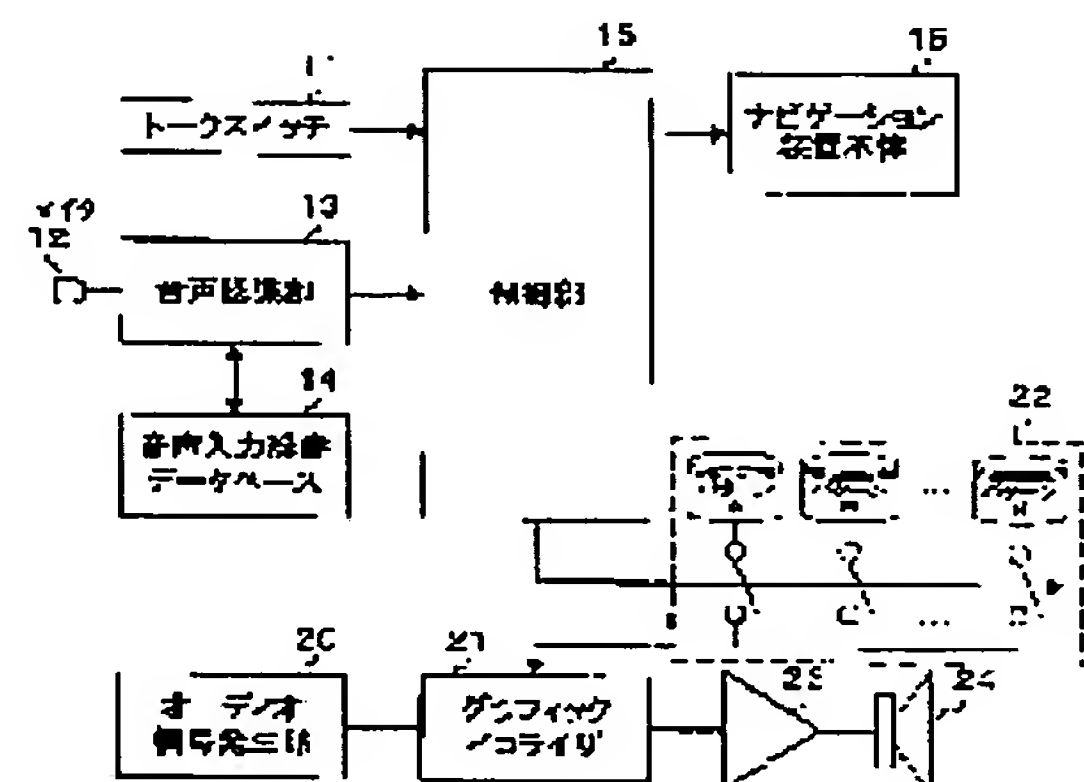
(72)Inventor : NAKADA KOICHI

## (54) METHOD AND DEVICE FOR SPEECH RECOGNITION

### (57)Abstract:

PROBLEM TO BE SOLVED: To provide a method and a device for speech recognition which can hold recognition precision in speech recognition excellent without lowering the output of an audio device so much.

SOLUTION: A control part 15 selects an (N)th pattern among a plurality of equalizer patterns stored in an equalizer pattern storage part 22 in speech recognition processing. Consequently, data of the (N)th pattern are inputted to a graphic equalizer 21, which attenuates the level of a signal component in a 100 to 5 kHz frequency band used for the speech recognition larger than signal components in other frequency bands.



## LEGAL STATUS

[Date of request for examination]

30.07.2004

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(51)Int.Cl. <sup>7</sup>	識別記号	F I	テ-マコ-ト*(参考)
G 1 0 L 15/20		B 6 0 R 11/02	B 2 F 0 2 9
B 6 0 R 11/02		G 0 1 C 21/00	A 3 D 0 2 0
G 0 1 C 21/00		G 0 6 F 3/16	3 2 0 H 5 D 0 1 5
G 1 0 L 15/00		G 1 0 L 3/02	3 0 1 C
15/28		3/00	5 5 1 Q
審査請求 未請求 請求項の数4 O L (全 6 頁) 最終頁に続く			

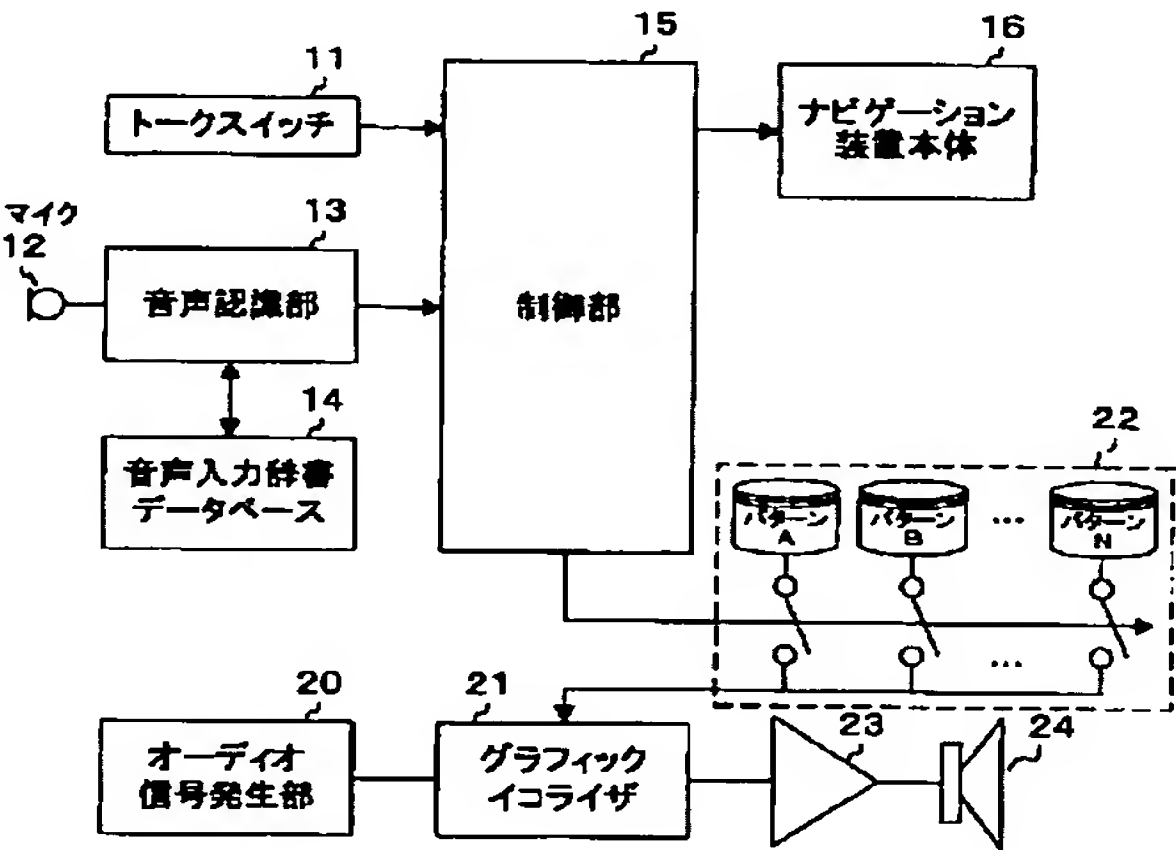
(21)出願番号	特願2001-307200(P2001-307200)	(71)出願人 000101732 アルパイン株式会社 東京都品川区西五反田1丁目1番8号
(22)出願日	平成13年10月3日(2001.10.3)	(72)発明者 中田 孝一 東京都品川区西五反田1丁目1番8号 ア ルパイン株式会社内
		(74)代理人 100091672 弁理士 岡本 啓三
		Fターム(参考) 2F029 AA02 AC18 3D020 BA02 BB01 BC01 BD03 BE04 5D015 EE04 KK01

(54)【発明の名称】 音声認識方法及び音声認識装置

(57)【要約】

【課題】 オーディオ装置の出力をそれほど低下させなくとも音声認識の際の認識精度を良好に維持できる音声認識方法及び音声認識装置を提供する。

【解決手段】 制御部15は、音声認識処理時にはイコライザパターン記憶部22に記憶されている複数のイコライザパターンのうちからN番目のパターンを選択する。これにより、N番目のパターンのデータがイコライザ21に入力され、グラフィックイコライザ21では音声認識に使用する100～5kHzの周波数帯域の信号成分のレベルを、他の周波数帯域の信号成分よりも大きく減衰させる。



## 【特許請求の範囲】

【請求項1】 音声認識処理を実行する際にオーディオ装置の音を自動的に低減する音声認識方法において、前記オーディオ装置から出力されるオーディオ信号のうち前記音声認識処理に使用する周波数帯域の信号成分の低減量を他の周波数帯域の信号成分の低減量よりも大きくすることを特徴とする音声認識方法。

【請求項2】 前記オーディオ装置から出力されるオーディオ信号のうち前記音声認識に使用する周波数帯域の信号成分を更に複数の周波数帯域に分割し、低域側の周波数帯域の信号成分の低減量を高域側の周波数帯域の信号成分の低減量よりも小さくすることを特徴とする請求項1に記載の音声認識方法。

【請求項3】 音声を入力し電気信号に変換する音声入力部と、前記音声入力部の出力を音声認識処理する音声認識部と、オーディオ信号を入力し、そのオーディオ信号の周波数帯域毎のレベルを個別に調整可能なレベル調整部と、前記レベル調整部による周波数帯域毎のレベル調整パターンを複数記憶するレベル調整パターン記憶部と、前記音声認識部及び前記レベル調整部を制御する制御部とを有し、前記制御部は、音声認識処理時に前記レベル調整パターン記憶部に記憶している複数のレベル調整パターンのうちから特定のレベル調整パターンを選択して、前記レベル調整部で音声認識時に使用する周波数帯域の信号成分を他の周波数帯域の信号成分よりも大きく減衰させることを特徴とする音声認識装置。

【請求項4】 前記特定のレベル調整パターンが、前記音声認識時に使用する周波数帯域の信号成分を更に複数の周波数帯域に分割し、低域側の周波数帯域の信号成分の低減量を高域側の周波数帯域の信号成分の低減量よりも小さくするものであることを特徴とする請求項3に記載の音声認識装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、音声認識処理を行うときにオーディオ装置の音をミュートする音声認識方法及び音声認識装置に関し、特に車載用オーディオ／ナビゲーションシステムに好適な音声認識方法及び音声認識装置に関する。

## 【0002】

【従来の技術】近年、音声認識技術が著しく発達し、種々の分野で使用されるようになった。例えば車載用ナビゲーション装置では、目的地の設定や画面の切替えなどの操作を音声でできるようにしたものも多い。

【0003】ところで、通常、車両にはCDの再生やラジオ放送の聴取が可能なオーディオ装置が搭載されており、音楽を聴きながら車両の運転を行うことが一般的に

行われている。しかし、音声認識によりナビゲーション装置を操作する場合は、周囲の音量が大きいと音声認識の精度が著しく低下してしまう。

【0004】このため、車載用ナビゲーション装置に内蔵された音声認識装置には、音声認識による操作を行うときはオーディオ装置の音量を自動的に無音化又は20dB程度下げる、いわゆるミュート機能が設けられている。

## 【0005】

10 【発明が解決しようとする課題】オーディオ装置の音量を低減することは、音声認識処理にとっては好都合であるものの、車両の乗員にとってはせっかく楽しんでいる音楽やラジオ放送が聞こえなくなったり、極めて小さな音量になってしまうので不満が生じる。特に、ニュースや交通情報などの有用な情報を聴き逃してしまった場合は、極めて大きな不満が生じる。

20 【0006】本発明の目的は、オーディオ装置の出力をそれほど低下させなくても音声認識の際の認識精度を良好に維持することができる音声認識方法及び音声認識装置を提供することである。

## 【0007】

【課題を解決するための手段】上記した課題は、音声認識処理を実行する際にオーディオ装置の音を自動的に低減する音声認識方法において、前記オーディオ装置から出力されるオーディオ信号のうち前記音声認識処理に使用する周波数帯域の信号成分の低減量を他の周波数帯域の信号成分の低減量よりも大きくすることを特徴とする音声認識方法により解決する。

30 【0008】本発明においては、オーディオ装置から出力されるオーディオ信号のうち、音声認識に使用する周波数帯域の信号成分の低減量を他の周波数帯域の信号成分の低減量よりも大きくするので、音声認識時のS/N比（ユーザの音声とそれ以外の音との比）が大きくなる。従って、オーディオ装置の出力をそれほど低下させなくても音声認識の認識精度を良好に維持することができる。

40 【0009】また、人間には、一般的に周囲の音が大きいと声が大きくなるという特性があるので、本発明の方法によれば、音声認識時のS/N比が従来に比べてより一層大きくなり、静寂な環境下での音声認識率に近づけることができる。

【0010】音声認識に使用する周波数帯域を更に複数の周波数帯域に分割し、低域側の周波数帯域の信号成分の低減量を高域側の周波数帯域の信号成分の低減量よりも小さくすることにより、音声認識率を良好に維持したまま、オーディオ装置から出力される音を聞き取りやすくすることができる。

50 【0011】上記した課題は、音声を入力し電気信号に変換する音声入力部と、前記音声入力部の出力を音声認識処理する音声認識部と、オーディオ信号を入力し、そ



のオーディオ信号の周波数帯域毎のレベルを個別に調整可能なレベル調整部と、前記レベル調整部による周波数帯域毎のレベル調整パターンを複数記憶するレベル調整パターン記憶部と、前記音声認識部及び前記レベル調整部を制御する制御部とを有し、前記制御部は、音声認識処理時に前記レベル調整パターン記憶部に記憶している複数のレベル調整パターンのうちから特定のレベル調整パターンを選択して、前記レベル調整部で音声認識時に使用する周波数帯域の信号成分を他の周波数帯域の信号成分よりも大きく減衰させることを特徴とする音声認識装置により解決する。

【0012】本発明においては、レベル調整パターン記憶部に周波数帯域毎のレベル調整パターンを複数記憶しておく。そして、音声認識処理時には、これらのパターンのうちから特定のレベル調整パターンを選択し、それによりレベル調整部で音声認識に使用する周波数帯域の信号成分を他の周波数帯域の信号成分よりも大きく減衰させる。このような構成とすることにより、オーディオ装置から出力される音をそれほど小さくしなくても、音声認識率の低下を回避することができる。また、本発明によれば、周波数帯域毎のレベルを個別に調整可能なレベル調整部を有するオーディオ装置の場合に、レベル調整パターン記憶部に上述した特定のレベル調整パターンを記憶しておくだけでよいので、装置コストの上昇を抑えることができる。

【0013】

【発明の実施の形態】以下、本発明について更に詳細に説明する。

【0014】図1～図3は、いずれも横軸に周波数を取り、縦軸に1Hz当りの長時間実効値(dB)をとって、人間の音声スペクトルを実測した例を示す図である(三浦種敏監、「新版 音声と聴覚」(電子情報通信学会、1980)、pp.330)。図1は女性5人の平均値と男性2人の平均値とを示し、図2は1人の男性による会話音声と音節のスペクトルを示し、図3は距離が3cm及び33.5cmでの女性音節スペクトルを示している。

【0015】これらの図から明らかなように、人間の音声には約100Hzから10kHz程度までの周波数成分が含まれ、100Hz～5kHzの周波数帯域の信号成分が大きな割合を占めている。このため、一般的な音声認識装置では、音声信号から約100Hz～5kHzの周波数帯域の信号成分を取り出して音声認識処理に用いている。

【0016】そこで、本発明においては、音声認識に使用されている周波数帯域とそれ以外の周波数帯域とでミュート量(低減量)を異なるものとする。これにより、音楽やニュース等を聞きながら、音声による機器の操作が可能になる。

【0017】(第1の実施の形態)図4は本発明の実施の形態の音声認識装置を備えた車載用オーディオ/ナビ

ゲーションシステムの構成を示す図である。

【0018】このシステムは、トークスイッチ11、マイク12、音声認識部13、音声入力辞書データベース14、制御部15、ナビゲーション装置本体16、オーディオ信号発生部20、グラフィックイコライザ21、イコライザパターン記憶部22、アンプ23及びスピーカー24により構成される。

【0019】トークスイッチ11は音声認識の有効/無効を切替えるスイッチである。音声認識によりナビゲーション装置を操作する場合はトークスイッチ11をオンにし、音声認識を使用しない場合はトークスイッチ11をオフにする。

【0020】マイク12は、運転席乗員の音声をよく收音できる場所に取り付けられる。音声認識部13は、マイク12から出力された音声信号を、音声入力辞書データベース14を参照して文字データ(又は、所定のコード)に変換し、制御部15に出力する。

【0021】制御部15は、音声認識部13から出力された文字データに応じた信号をナビゲーション装置本体16に伝達する。ナビゲーション装置本体16は、制御部15から伝達された信号に応じた処理を実行する。

【0022】オーディオ信号発生部20は、オーディオ信号を出力するCDプレーヤ及びラジオチューナ等となる。グラフィックイコライザ21は、オーディオ信号発生部20から出力されたオーディオ信号を複数の周波数帯域に分割し、各周波数帯域毎にレベル(音量)を調整した後、各周波数帯域の信号成分を再び合成して1つのオーディオ信号とする。イコライザパターン記憶部22には、複数のイコライザパターン(レベル調整パターン)が記憶されており、グラフィックイコライザ21は、イコライザパターン記憶部22に記憶されている複数のイコライザパターンのうちから制御部15により選択されたパターンのデータが入力され、そのデータに応じて各周波数帯域毎の信号成分のレベルを調整する。イコライザパターンには、例えば、低音域及び高音域の周波数成分のレベルを高くしてメリハリのある音楽を再生するパターンや、中音域の周波数成分のレベルを高くして人の声を強調するようなパターンや、低音域から高音域までフラットなレベルにして自然な音質を得るパターンなどがある。これらのイコライザパターンは、ユーザの好みに応じて適宜選択される。

【0023】但し、本実施の形態では、N番目のイコライザパターンとして、音声認識時に使用されるパターンが記憶されている。このイコライザパターンは、例えば下記表1に示すように、20Hz～100Hzの周波数帯域及び5kHz～20kHzの周波数帯域では-5dBのミュート量(減衰量)が設定され、音声認識に使用される帯域である100Hz～5kHzの帯域では-20dBのミュート量が設定されている。なお、表1に示す数値は一例であり、本発明において、音声認識に使用

する周波数帯域の範囲、及び各周波数帯域毎のミュート量はこれに限定するものではない。

〔0024〕

【表1】

周波数帯域	ミュート量
20Hz～100Hz	-5dB
100Hz～5kHz	-20dB
5kHz～20kHz	-5dB

〔0025〕以下、本実施の形態のオーディオ／ナビゲーションシステムの動作について説明する。

〔0026〕トークスイッチ11がオフのとき、すなわち音声認識による操作を行わないときは、制御部15はイコライザパターン記憶部22に記憶されている複数のイコライザパターンのうちからユーザの指示に応じたパターンを選択する。これにより、グラフィックイコライザ21には選択されたパターンに応じたデータが入力され、オーディオ信号発生部20から出力されたオーディオ信号が周波数帯域毎にレベル調整される。そして、グラフィックイコライザ21から出力されたオーディオ信号はアンプ23で増幅され、スピーカ24に供給される。これにより、スピーカ24からユーザの好みに応じた音質で音楽等が再生される。

〔0027〕ユーザが音声認識によりナビゲーション装置を操作する場合、トークスイッチ11をオンにする。これにより、制御部15はイコライザパターン記憶部22に記憶されている複数のイコライザパターンからN番目のパターンを選択する。

〔0028〕このN番目のパターンのデータが入力されると、グラフィックイコライザ21は、上記の表1に示すように、オーディオ信号発生部20から出力されるオーディオ信号のうち100Hz～5kHzの周波数帯域の信号成分を-20dB減衰させ、20Hz～100Hz及び5kHz～20kHzの周波数帯域の信号成分を-5dB減衰させる。その結果、音声認識に使用する周波数帯域の信号成分が大幅に減衰されるものの、それ以外の周波数帯域の信号成分の減衰量が少ないので、聴感上はスピーカ24から出力される音の音量がそれほど大きく下がったようには感じない。

〔0029〕この状態でユーザが発音すると、マイク12から音声に応じた電気信号が出力されて、音声認識部13に入力される。音声認識部13は音声入力辞書データベース14を参照しながら入力された電気信号を文字データに変換して出力する。

〔0030〕制御部15は、音声認識部13から出力された文字データに応じた信号をナビゲーション装置本体16に伝達し、ナビゲーション装置本体16では制御部

15から伝達された信号に応じた処理を実行する。

〔0031〕上述したように、本実施の形態では、音声認識によりナビゲーション装置を操作するときには、オーディオ信号発生部20から出力されたオーディオ信号のうち音声認識に使用する100Hz～5kHzの周波数帯域の信号成分を大きく減衰させるので、スピーカ24から出力される音が音声認識時の障害となることなく、認識率の低下が回避される。

〔0032〕また、本実施の形態では、オーディオ信号発生部20から出力されたオーディオ信号のうち20Hz～100Hz及び5kHz～20kHzの周波数帯域の信号成分の減衰量が小さいので、聴感上はスピーカ24から出力される音の音量がそれほど小さくなったようには感じない。従って、音楽やニュース等が聞こえないというような不満が解消される。

〔0033〕ところで、通常人間は、周囲の音が大きいと大きな声を出すようになる。本実施の形態では、音声認識に使用する周波数帯域以外のミュート量が少ないので、従来のようにオーディオ装置から出力される音を全体的に小さくする場合に比べて、ユーザの声が大きくなる。このため、音声認識時のSN比、すなわちユーザの音声とそれ以外の音との比が大きくなり、オーディオ装置を使用している環境下での音声認識率の低下をより確実に回避できるという効果もある。

〔0034〕更にまた、本実施の形態においては、グラフィックイコライザを有するオーディオ装置の場合、表1に示すような音声認識処理用のイコライザパターンをイコライザパターン記憶部に記憶しておくだけでよいので、製品コストの上昇を抑えることができるという利点もある。

〔0035〕（第2の実施の形態）以下、本発明の第2の実施の形態について説明する。本実施の形態も、本発明を車載用オーディオ／ナビゲーションシステムに適用した例を示す。

〔0036〕本実施の形態が第1の実施の形態と異なる点は、音声認識時に選択されるイコライザパターンが異なることにあり、車載用オーディオ／ナビゲーションシステムの基本的な構成は第1の実施の形態と同様であるので、本実施の形態においても第4図を参照して説明する。

〔0037〕図1～図3に示すように、人間の声は、約100Hzから10kHz程度までの周波数成分を含んでいるが、更に詳細に調べると、低周波成分を多く含むそのレベルが大きい母音と、高周波成分を多く含むそのレベルが小さい子音とに分類することができる。また、音声認識処理ではSN比、すなわちユーザの声とそれ以外の音との比が大きいことが重要である。

〔0038〕上記のことから、母音の周波数帯域では、子音の周波数帯域に比べてミュート量が少なくてもよいことがわかる。そこで、本実施の形態においては、下記

表2に示すように、音声認識時の各周波数帯毎のミュート量を設定する。すなわち、子音の周波数帯域である1kHz～2kHz及び2kHz～5kHzのミュート量をそれぞれ-18dB及び-21dBと比較的大きくしているのに対し、母音の周波数帯域である100Hz～1kHzのミュート量を-15dBと比較的小さくしている。なお、表2に示す数値は一例であり、本発明において、周波数帯域の分割数、各周波数帯域の範囲、及び各周波数帯域毎のミュート量はこれに限定されるものではない。

【0039】

【表2】

周波数帯域	ミュート量
20Hz～100Hz	-5dB
100Hz～1kHz	-15dB
1kHz～2kHz	-18dB
2kHz～5kHz	-21dB
5kHz～20kHz	-5dB

【0040】本実施の形態によれば、音声認識に使用される周波数帯域を更に低域側の母音の周波数帯域と高域側の子音の周波数帯域とに分割し、母音の周波数帯域では子音の周波数帯域に比べてミュート量を小さくしている。これにより、第1の実施の形態に比べて、スピーカ-24から出力される音楽やニュース等の情報番組をより自然に近い音質で聞くことができる。従って、本実施の形態においては、第1の実施の形態に比べて、ニュースや交通情報などの聴取が妨げられる程度がより一層少なくなるという利点がある。

【0041】なお、上述した第1及び第2の実施の形態ではいずれも音声認識に使用する周波数帯域を100Hz～5kHzとしたが、例えば300Hz～4.8kHzの周波数帯域を音声認識に使用する音声認識装置の場合は、300Hz～4.8kHzの周波数帯域のミュート量を大きくし、それ以外の周波数帯域のミュート量を小さくすればよい。

【0042】また、上述した第1及び第2の実施の形態はいずれも本発明を車載用オーディオ／ナビゲーションシステムに適用した場合について説明したが、これにより本発明の適用範囲が車両に搭載される機器の音声認識

装置に限定されるものではない。更に、第1及び第2の実施の形態はいずれもレベル調整部としてグラフィックイコライザを用いた場合について説明したが、グラフィックイコライザに替えてバンドパスフィルタやパラメトリックイコライザ等を使用してもよい。

【0043】

【発明の効果】以上説明したように、本発明によれば、オーディオ装置から出力されるオーディオ信号のうち、音声認識処理に使用する周波数帯域の信号成分の減衰量を他の周波数帯域の信号成分の減衰量よりも大きくするので、オーディオ装置の出力をそれほど小さくしなくても、良好な音声認識率を得ることができる。

【0044】また、音声認識に使用する周波数帯域の信号成分を更に複数の周波数帯域に分割し、低域側の周波数帯域の信号成分の減衰量を高域側の周波数帯域の信号成分の減衰量に比べて小さくすることにより、音声認識率を良好に維持したまま、オーディオ装置から出力される音を聞き取りやすくすることができる。

【図面の簡単な説明】

【図1】図1は人間の音声スペクトルを実測した例を示す図であり、女性5人の平均値と男性2人の平均値とを示している。

【図2】図2は人間の音声スペクトルを実測した例を示す図であり、1人の男性による会話音声と音節のスペクトルを示している。

【図3】図3は人間の音声スペクトルを実測した例を示す図であり、距離が3cm及び33.5cmでの女性音節スペクトルを示している。

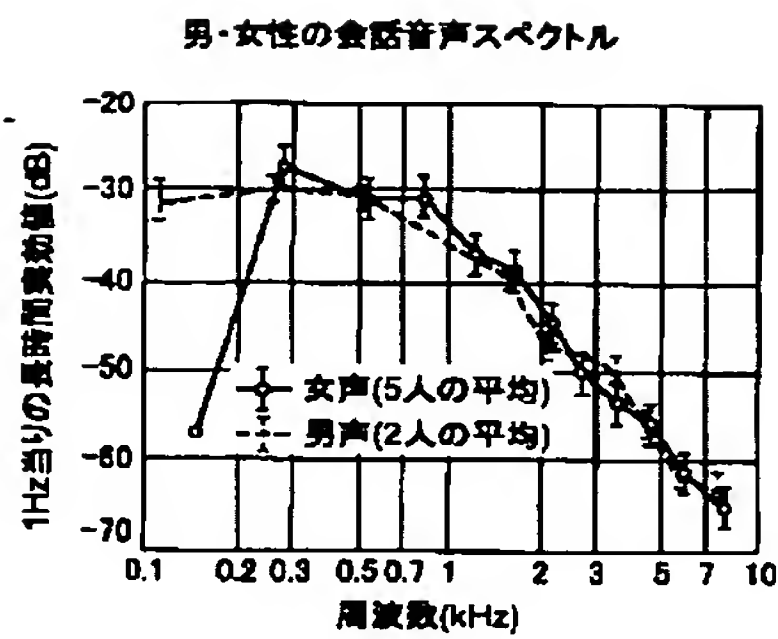
【図4】図4は本発明の実施の形態の音声認識装置を備えた車載用オーディオ／ナビゲーションシステムの構成を示す図である。

【符号の説明】

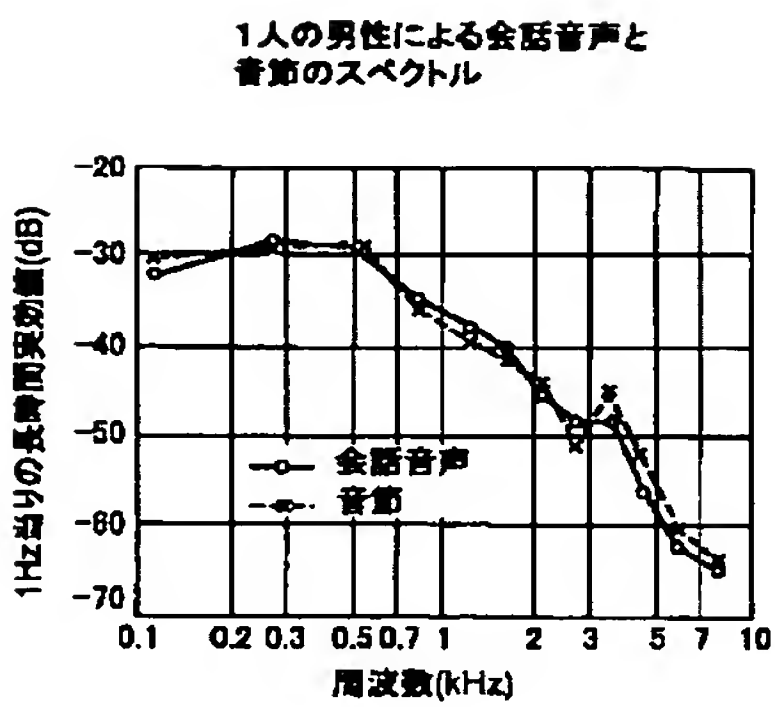
- 11…トークスイッチ、
- 12…マイク、
- 13…音声認識部、
- 14…音声入力辞書データベース、
- 15…制御部、
- 16…ナビゲーション装置本体、
- 20…オーディオ信号発生部、
- 21…グラフィックイコライザ、
- 22…イコライザパターン記憶部、
- 23…アンプ、
- 24…スピーカ。



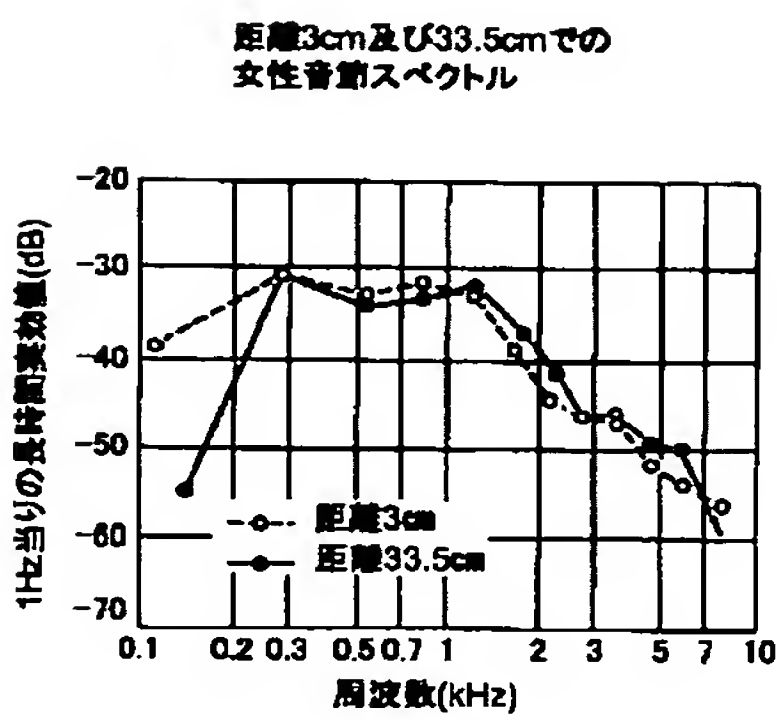
【図1】



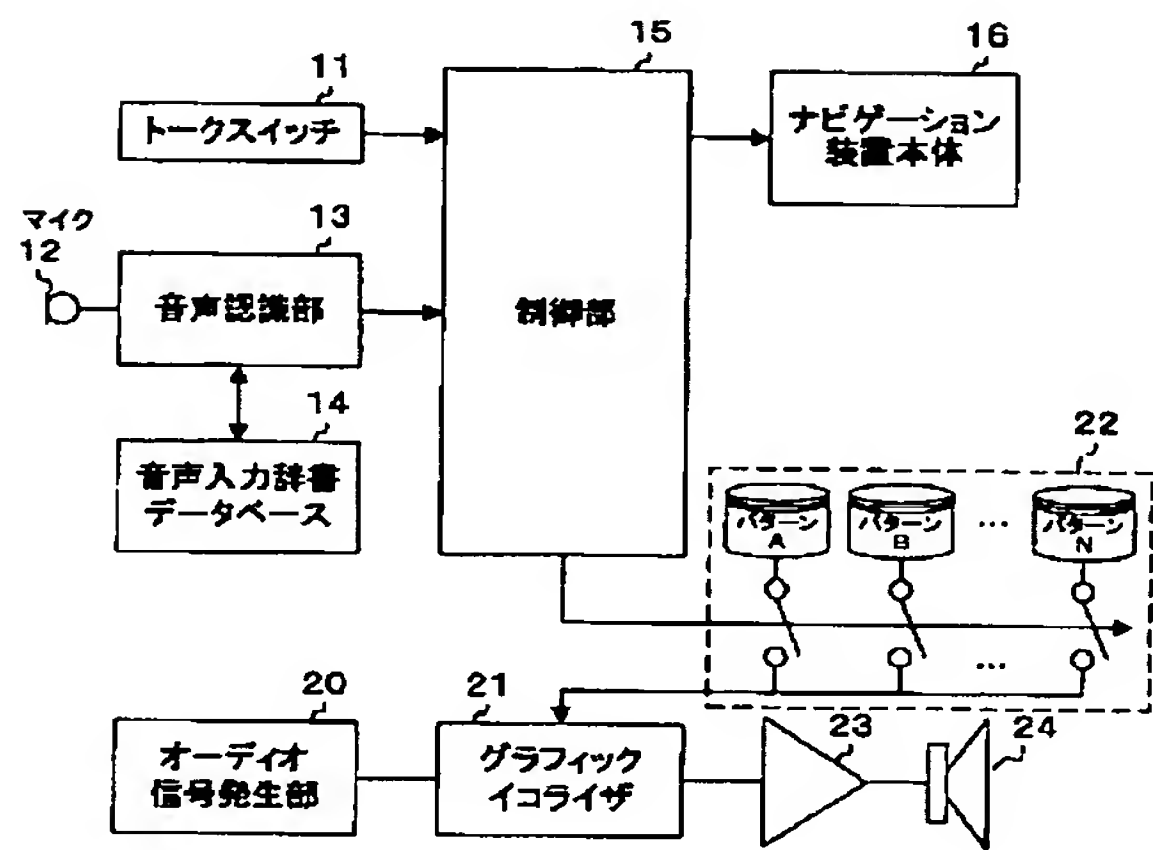
【図2】



【図3】



【図4】



フロントページの続き